

Quantitative Structure-Activity Relationship of Insecticidal Activity of Benzyl Ether Diamidine Derivatives

Mengting Zhai, Yan Chen*, Jing Li, Jun Zhou

School of Chemistry and Chemical Engineering, Xuzhou University of Technology, Xuzhou 221111

*Corresponding author: chenyan681110@126.com

Abstract: The molecular electronegativity distance vector (MEDV-13) was used to describe the molecular structure of benzyl ether diamidine derivatives in this paper, Based on MEDV-13, The three-parameter (M_3 , M_{15} , M_{47}) QSAR model of insecticidal activity (pIC_{50}) for 60 benzyl ether diamidine derivatives was constructed by leaps-and-bounds regression (LBR). The traditional correlation coefficient (R) and the cross-validation correlation coefficient (R_{CV}) were 0.975 and 0.971, respectively. The robustness of the regression model was validated by Jackknife method, the correlation coefficient R were between 0.971 and 0.983. Meanwhile, the independent variables in the model were tested to be no autocorrelation. The regression results indicate that the model has good robust and predictive capabilities. The research would provide theoretical guidance for the development of new generation of anti African trypanosomiasis drugs with efficiency and low toxicity.

1. Introduction

Human African trypanosomiasis, also known as African sleeping sickness, is a zoonotic parasitic disease, the disease can cause nerve dysfunction, blurred consciousness, poor physical coordination and other symptoms, cause permanent nerve injury[1,2]. This disease has spread all over the world, and has been controlled by the unremitting efforts of medical scientists all over the world. However, due to the slow development of new drugs, the current treatment still has some problems.

A series of benzyl ether diamidine derivatives which show good insecticidal activity and low toxicity have been synthesized by Patrick[3], They are expected to be a new generation of anti-African trypanosomiasis drugs. However, the structure-activity relationship and the mechanism of action of these drugs remain largely unknown. Quantitative structure activity relationship analysis (QSAR)[4-6] has played an important role in the design and screening of highly effective drugs and the mechanism of action of drugs.

On the basis of previous work[7-8], molecular electronegativity distance vector (M_i) of 60 benzyl ether diamidine derivatives were calculated using the MATLAB programs[9-10], and by optimal subset regression method the three best variables were screened to establish the QSAR model. The model was tested to be highly reliable and has favorable predictive ability.

2. Materials and methods

2.1 Molecular Structure of Benzyl Ether Diamidine Derivatives

The matrix structure of benzyl ether diamidine derivatives was shown in fig. 1. All 60 molecular structures were shown in Table 1. Experimental values of their insecticidal activity (pIC_{50} , the minus



paired-domination number of IC_{50} (50% inhibitory concentration)) were obtained from the literature[11]

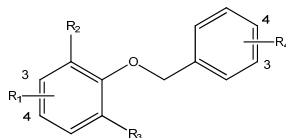


Fig.1 The matrix structure of benzyl ether diamidine derivatives

Table 1. The insecticidal activity of 60 benzyl ether diamidine derivatives

Comp	R_1	R_2	R_3	R_4	pIC_{50}		
					Exp.	Cal.	Err.
1	4-Am	H	H	H	5.735	5.313	-0.422
2	4-Am	I	H	4-NO ₂	5.721	5.881	0.160
3	4-Am	I	I	3-NO ₂	5.575	6.084	0.509
4	3-Am	H	H	H	5.031	5.336	0.305
5	3-Am	H	H	3-NO ₂	6.318	5.535	-0.783
6	H	H	H	4-Am	5.431	5.320	-0.111
7	4-NO ₂	H	H	4-Am	5.356	5.535	0.179
8	4-Cl	H	H	4-Am	5.567	5.520	-0.047
9	4-iPr	H	H	4-Am	5.889	6.322	0.433
10	4-OMe	H	H	4-Am	5.403	5.435	0.032
11	4-NO ₂	H	H	4-Am	5.523	5.471	-0.052
12	3-NO ₂	H	H	4-Am	5.674	5.519	-0.155
13	3-NO ₂	H	H	3-Am	5.138	5.543	0.405
14	4-Am	H	H	4-Am	7.041	7.459	0.418
15	4-iPrAm	H	H	4-iPrAm	5.870	5.894	0.024
16	4-Am	Cl	H	4-Am	7.658	7.672	0.014
17	4-Am	Br	H	4-Am	7.796	7.779	-0.017
18	4-AmOH	Br	H	4-AmOH	4.620	4.500	-0.120
19	4-AmOMe	Br	H	4-AmOMe	4.726	4.433	-0.293
20	4-Am	I	H	4-Am	7.770	7.829	0.059
21	4-AmOH	I	H	4-AmOH	4.588	4.546	-0.042
22	4-AmOMe	I	H	4-AmOMe	4.298	4.477	0.179
23	4-Am	Cl	Cl	4-Am	7.620	7.785	0.165
24	4-Am	Br	Br	4-Am	7.770	7.979	0.209
25	4-Am	I	I	4-Am	6.914	8.071	1.157
26	4-Am	OMe	OMe	4-Am	6.959	7.380	0.421
27	4-AmOH	OMe	OMe	4-AmOH	3.793	3.962	0.169
28	4-AmOMe	OMe	OMe	4-AmOMe	3.889	3.859	-0.030
29	4-Am	H	H	3-Am	7.620	7.515	-0.105
30	4-Am	NO ₂	H	3-Am	7.469	7.554	0.085
31	4-Am	Cl	H	3-Am	8.155	7.730	-0.425
32	4-AmOH	Cl	H	4-AmOH	4.308	4.399	0.091
33	4-AmOMe	Cl	H	4-AmOMe	4.583	4.331	-0.252
34	4-Am	Br	H	3-Am	8.301	7.837	-0.464
35	4-AmOH	Br	H	4-AmOH	4.833	4.497	-0.336
36	4-AmOMe	Br	H	4-AmOMe	4.102	4.423	0.321
37	4-Am	I	H	3-Am	8.301	7.888	-0.413

38	4-AmOH	I	H	4-AmOH	4.299	4.543	0.244
39	4-AmOMe	I	H	4-AmOMe	4.460	4.467	0.007
40	4-Am	OMe	H	3-Am	7.721	7.572	-0.149
41	4-Am	Cl	Cl	3-Am	8.000	7.844	-0.156
42	4-AmOH	Cl	Cl	4-AmOH	4.578	4.454	-0.124
43	4-AmOMe	Cl	Cl	4-AmOMe	4.221	4.362	0.141
44	4-Am	Br	Br	3-Am	8.398	8.039	-0.359
45	4-AmOH	Br	Br	4-AmOH	5.162	4.629	-0.533
46	4-AmOMe	Br	Br	4-AmOMe	4.157	4.523	0.366
47	4-Am	I	I	3-Am	8.523	8.131	-0.392
48	4-AmOH	I	I	4-AmOH	4.866	4.713	-0.153
49	4-AmOMe	I	I	4-AmOMe	4.580	4.601	0.021
50	4-Am	Br	OMe	3-Am	8.301	7.719	-0.582
51	4-AmOH	Br	OMe	4-AmOH	4.056	4.270	0.214
52	4-AmOMe	Br	OMe	4-AmOMe	4.089	4.161	0.072
53	4-Am	I	OMe	3-Am	8.523	7.757	-0.766
54	4-AmOH	I	OMe	4-AmOH	4.412	4.304	-0.108
55	4-AmOMe	I	OCH ₃	4-AmOMe	4.124	4.191	0.067
56	4-Am	OMe	OMe	3-Am	6.738	7.440	0.702
57	4-AmOH	OMe	OMe	4-AmOH	3.804	3.953	0.149
58	4-AmOMe	OMe	OMe	4-AmOMe	3.963	3.841	-0.122
59	3-Am	H	H	4-Am	7.553	7.513	-0.040
60	3-Am	H	H	3-Am	7.013	7.576	0.563

2.2 Calculation of Molecular Electronegativity Distance Vector(MEDV-13)

The molecular electronegativity distance vector based on 13 atomic types, called MEDV-13, is a descriptor for predicting the biological activities of molecules based on the quantitative structure-activity relations (QSAR). The MEDV-13 uses a modified electrotopological state (E-state) index to substitute for the relative electronegativity (q) of non-hydrogen atoms in the molecule of interest in the MEDV and a topological distance for the relative distance (d) in the MEDV[12].

$$M_i = M_{kl} = \sum_{i \in k, j \in l} \frac{q_i q_j}{d_{ij}^2} \quad (1)$$

$$(k, l = 1, 2, 3, \dots, 13; 1 \leq k)$$

Where q_i is the E-state index for atom i , d_{ij} is the graph distance between two atoms, atom i and j . Using Chem3D Ultra 9.0, the molecular structures of 60 benzyl ether diamidine derivatives were built, and the molecular electronegativity distance vectors were calculated by MATLAB. The MEDV-13 includes at best 91 descriptors. According to the principle of statistics, the number of variables is less than 5% of the independent variable, and its contribution to the dependent variable can be neglected. Therefore, the independent variables which number was less than 3 ($60 \times 5\%$) were removed, and the remaining 41 topological indexes were used to characterize the molecular structure of 60 benzyl ether diamidine derivatives

3. Results and Discussion

3.1 Analysis of Statistical Regression.

Using the remaining 41 topological indexes of MEDV as independent variables, and insecticidal activity (pIC_{50}) as dependent variables. We had chosen the best variables of relativity with insecticidal activity by MINITAB 14 software, and then built up the mathematical model between MEDV-13 indexes and pIC_{50} . The best subset regression analysis results were shown in Table 2.

Table 2 Results of the topological index and insecticidal activity with the leaps-and-bounds regression

NO.	R	R^2	R_{adj}^2	S	F	AIC	FIT	variable
1	0.951	0.905	0.903	0.485	549.7	0.240	9.209	M_{47}
2	0.962	0.926	0.924	0.431	357.3	0.194	11.140	M_{47} 、 M_3
3	0.975	0.950	0.948	0.356	357.7	0.138	15.420	M_{47} 、 M_3 、 M_{15}
4	0.976	0.952	0.948	0.354	272.3	0.141	14.353	M_{47} 、 M_3 、 M_{15} 、 M_2

In Table 2, R is the traditional correlation coefficient, R^2 is the determination coefficient, R_{adj}^2 is the square of adjusted correlation coefficient, S is standard error, and F is the test value of Fischer. We applied the Akaike's information criterion (AIC ; Eq. 2) and Kubinyi function (FIT ; Eq. 3) [13-14] to determine if a variable should be included in the model. That is to say, if the Akaike's information criterion decreases in value when adding an additional variable and the Kubinyi function increases in value, the introduction of this new variable is justified.

$$AIC = R_{SS} \times \frac{f+b}{(f-b)^2} \quad (2)$$

$$FIT = \frac{R^2(f-b-1)}{(f+b^2)(1-R^2)} \quad (3)$$

Where R_{SS} is the residual sum of squares, f is the number of compounds included in the model, b is the number of variables included in the model, and R^2 is the determination coefficient.

As can be seen in Table 2, AIC value gradually decreased and FIT value gradually increased, the turning point was 0.138 and 15.420, respectively, Corresponding multiple regression equations were shown as follows:

$$pIC_{50} = 2.828 + 0.480M_{47} - 0.202M_3 + 0.05 M_{15} \quad (4)$$

$$n = 60 \quad R = 0.975 \quad S = 0.356 \quad F = 357.7$$

Calculated values of pIC_{50} given by the formula (4) were listed in Table 1, which were in agreement with the corresponding experimental values. Fig 2 was the plot of predicted against experimental values of the insecticidal activity (pIC_{50})

3.2 Stability tests of the model

3.2.1 LOO cross-validation of the model

By LOO cross-validation correlation coefficient (R_{cv}^2) was 0.943 and slightly smaller than 0.950; and $R_{cv}^2 > 0.5$. These show that the stability and prediction ability of the model are ideal.

3.2.2 Autocorrelation test of the independent variables

The autocorrelation of the independent variables of the model was tested using variance inflation factors (VIF , Eq. 5)

$$VIF = \frac{1}{1-R^2} \quad (5)$$

Where R^2 is the determination coefficient of one independent variable with the remaining independent variables. If $VIF = 1$, there was no autocorrelation between independent variables, if $VIF < 5$, the autocorrelation is weak and the model is stable, and if $VIF \geq 5$, there is obvious autocorrelation and the model can not be used to estimate and forecast. The VIF value of M_3 , M_{15} , M_{47} in model (4) are 1.117, 1.502 and 1.368, respectively. The result demonstrates that there is no autocorrelation between independent variables in model (4).

3.2.3 Jackknifed test of the model

To test whether there was any "abnormal value" in model (4), we carried through the stability test using Jackknifed method [16]. Every time we eliminated the compounds whose serial number contains 1, 2, 3 ... 0 on unit order in the sample, and then established the model with the rest compounds. Repeat 10 times and 10 correlation coefficients (R) and standard error (S) were got (Table 2),. From

Table 2, all *R* values and *S* values were relatively close. It suggests that the model (4) has good robustness.

Table 3. The robustness test of model (4)

Delete compounds	1	2	3	4	5	6	7	8	9	0
<i>R</i>	0.974	0.973	0.976	0.974	0.983	0.977	0.971	0.973	0.974	0.974
<i>S</i>	0.364	0.374	0.345	0.358	0.301	0.348	0.364	0.374	0.368	0.358

4. Summary

(1) A molecular electronegativity distance vector (MEDV-13) contain the topological, geometrical and electrical characteristics of the compounds, so there is a good correlation between the insecticidal activity and MEDV-13

(2) M_3 , M_{15} , M_{47} are the interaction between the first class atom (CH_3 -) and third class atom ($>\text{CH}$ -), the second class atom ($-\text{CH}_2$ -) and the third class atom, the fifth class atom ($-\text{NH}_2$), and fifth class atom, so the main structural fragments influencing the insecticidal activity of benzyl ether diamidine derivatives are $-\text{CH}_3$, $-\text{CH}_2$ -, $>\text{CH}$ - and $-\text{NH}_2$.

(3) By LOO cross-validation, Autocorrelation test of the independent variables and Jackknifed test, established mathematical model by molecular electronegativity distance vector has a good robustness and prediction ability.

This method can provide some theoretical insights into the design of this series of medicine with high insecticidal activity.

Acknowledgement

This work was supported by the National Natural Science Found of China (No. 21272095)

References

- [1] P P Simarro, G Cecchi, M Paone, et al.: Int. J. Health. Geogr.2010, 9 (1): 57-75
- [2] R Brun, J Blum, F Chappuis, C Burri: Lancet. 2010, 375 (9709): 148-159
- [3] D A Patrick, S A Bakunov, S M Bakunova, et al: Eur. J. Med. Chem. 2013, 67: 310-324
- [4] J. Liu, X. Guan, J. Q. Peng: Acta. Chim. Sin. 2012,70:83-91 (In Chinese)
- [5] C. Y. W. Nina, V. Raghuraman, M. B. Robert: Reg. Toxic. Pharm. 2011,59: 215-226
- [6] Y. L. Liu, Y. T.Li, B.Z.Shi: Sci. Sin. Chim. 2013,43:198-208 (In Chinese)
- [7] Y Chen .Bull Sci Technol, 2013,29: 16-19 (In Chinese)
- [8]Y Chen. Food Sci. 2011, 32: 274-277 (In Chinese)
- [9] Q. N. Hu, Y. Z. Liang, Y. L. Wang Comp. App. Chem. 2003,20: 386-390 (In Chinese)
- [10] T. Zhang, Y.Z. Liang, C.X. Zhao. Chin. J. Anal. Chem.2006,34: 1607-1610 (In Chinese)
- [11] D A Patrick, S A Bakunov, S M Bakunova, et al. Eur. J. Med. Chem. 2013, 67: 310-324
- [12] S S Liu, C S Yin, Z Y, Li, et al. J. Chem. Inf. Comput. Sci. 2001, 41, 321-329
- [13] L. S. Urrea, M. P. Gonza'lez, M Teijeira. Bioorg. Med. Chem. 2007, 15, 3565-3571
- [14] L. S. Urrea, M. P. Gonza'lez, M. Teijeira. Bioorg. Med. Chem. 2006, 14, 7347-7358
- [15] C.J.Feng. Acta Physico- Chimica Sinica, 2010, 26: 193-198(in Chinese)
- [16] W S Dietrich, N D Dreyer, C Hansch. J Med Chem, 1980, 23: 1201- 1205